UNIVERSITY OF TECHNOLOGY SYDNEY
Faculty of Engineering and Information Technology

CONSENSUS-BASED DATA MANAGEMENT WITHIN FOG COMPUTING FOR THE INTERNET OF THINGS

by

Firas Qais Mahammed Saleh Al-Doghman

A Thesis Submitted
in Partial Fulfillment of the
Requirements for the Degree

Doctor of Philosophy

Sydney, Australia

2019

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Firas Qais Mohammed Saleh Al-Doghman declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Electrical and Data Engineering/Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Production Note:

Signature: Signature removed prior to publication.

Date: 4/12/2019

ABSTRACT

CONSENSUS-BASED DATA MANAGEMENT WITHIN FOG COMPUTING FOR THE INTERNET OF THINGS

by

Firas Qais Mahammed Saleh Al-Doghman

The Internet of Things (IoT) infrastructure forms a gigantic network of interconnected and interacting devices. This infrastructure involves a new generation of service delivery models, more advanced data management and policy schemes, sophisticated data analytics tools, and effective decision making applications. IoT technology brings automation to a new level wherein nodes can communicate and make autonomous decisions in the absence of human interventions. IoT enabled solutions generate and process enormous volumes of heterogeneous data exchanged among billions of nodes. This results in Big Data congestion, data management, storage issues and various inefficiencies. Fog Computing aims at solving the issues with data management as it includes intelligent computational components and storage closer to the data sources. Often, an IoT-enabled infrastructure is shared among many users with various requirements. Sharing resources, sharing operational costs and collective decision making (consensus) among many stakeholders is frequently neglected. This research addresses an essential requirement for adaptive, autonomous and consensus-based Fog computational solutions which are able to support distributed and in-network schemes and policies. These network schemes and policies need to meet the requirements of many users. In this work, innovative consensus-based computational solutions are investigated. These proposed solutions aim to correlate and organise data for effective management and decision making in Fog. Instead of individual decision making, the algorithms aim to aggregate several decisions into a consensus decision representing a collective agreement, benefiting from the individuals variant knowledge and meeting multiple stakeholders requirements. In order to validate the proposed solutions, hybrid research methodology is involved that includes the design of a test-bed and the execution of several experiments. In order to investigate the effectiveness of the paradigm, three experiments were designed and validated. Real-life sensor data and synthetic statistical data was collected, processed and analysed. Bayesian Machine Learning models and Analytics were used to consolidate the design and evaluate the performance of the algorithms. In the Fog environment, the first scenario tests the Aggregation by Distribution algorithm. The solution contribute in achieving a notable efficiency of data delivery obtained with a minimal loss in precision. The second scenario validates the merits of the approach in predicting the activities of high mobility IoT applications. The third scenario tests the applications related to smart home IoT. All proposed Consensus algorithms use statistical analysis to support effective decision making in Fog and enable data aggregation for optimal storage, data transmission, processing and analytics. The final results of all experiments showed that all the implemented consensus approaches surpass the individual ones in different performance terms. Formal results also showed that the paradigm is a good fit in many IoT environments and can be suitable for different scenarios when applying data analysis to correlate data with the design. Finally, the design demonstrates that Fog Computing can compete with Cloud Computing in terms of accuracy with an added preference of locality.

Dissertation directed by Dr. Zenon Chaczko and Dr. Wayne Brooke
School of Electrical and Data Engineering

Dedication

I dedicate my dissertation work to my family, my supervisor and many friends. A special feeling of gratitude to my loving parents whose words of encouragement and push for tenacity ring in my ears and whose a look of happiness in their faces is my inspiration. In the same time, my beloved children who have never left my mind and who I work to make them be proud of me. My Supervisors, who guided and supported me along the way with a beautiful spirit and care. My friends, for always being there for me and help me with the best they know. Thank you all for the love, compassion and support.

Acknowledgements

I would like to thank my supervisors, Dr. Zenon Chaczko and Dr. Wayne Brooke, who guided and supported my work along the way and helped me get results of better quality. I am also grateful to my friend, Abdelwahed Khamiss, for his help and support in overcoming numerous obstacles I have been facing through my research. I am also grateful to my friends Mahmoud Gamal and Alina Rakhi who helped me in several occasions to improve my research directions and writing skills. I would like to thank my fellow doctoral students for their feedback, cooperation and of course friendship. I would like to thank my friends for accepting nothing less than excellence from me. Last but not the least, I would like to thank my family: my parents, my children and my sisters for supporting me spiritually throughout writing this thesis and my life in general.

Firas Al-Doghman
Sydney, Australia, 2019.

List of Publications

Journal Papers

J-2. **F. Al-Doghman**, Z. Chaczko, and W. Brooke "Data Science for the IoT Fog Computing" *IEEE Internet of Things Journal*, in communication.

Conference Papers

C-1. **F. Al-Doghman**, Z. Chaczko and W. Brooks 2018, "Adaptive Consensus-based Aggregation for Edge Computing *2018 26th International Conference on Systems Engineering (ICSEng).*

C-2. A. Ajayan, **F. Al-Doghman** and Z. Chaczko 2018, "Visualizing Multimodal Big Data Anomaly Patterns in Higher-Order Feature Spaces *2018 26th International Conference on Systems Engineering (ICSEng).*

C-3. **F. Al-Doghman**, Z. Chaczko and J. Jiang 2017, "A review of aggregation algorithms for the internet of things *2017 25th International Conference on Systems Engineering (ICSEng).*

C-4. J. Jiang, Z. Chaczko, **F. Al-Doghman** and W. Narantaka 2017, "New LQR Protocols with Intrusion Detection Schemes for IOT Security *2017 25th International Conference on Systems Engineering (ICSEng).*

C-5. **F. Al-Doghman**, Z. Chaczko, A. Ajayan and R. Klempous 2016, "A review on Fog Computing technology *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC).*

C-6. **F. Al-Doghman**, Z. Chaczko, W. Brooks D., Hoang, and L. Carrion Gordon 2019, "Social Consensus-inspired Aggregation for Edge Computing *CSNet 2019 3rd Cyber Security in Networking conference.*

Book Chapter

C-1. **F. Al-Doghman**, Z. Chaczko, and A. Ajayan 2019, "Policy-based Consensus Data Aggregation for the Internet of Things *Smart Innovations in Engineering and Technology*, Topics in Intelligent Engineering and Informatics series.

C-2. A. Ajayan,**F. Al-Doghman** and Z. Chaczko 2019, "Tensor decompositions in multimodal Big Data: studying multiway behavioral patterns *Smart Innovations in Engineering and Technology*, Topics in Intelligent Engineering and Informatics series.

Contents

Certificate ii

Abstract iii

Dedication v

Acknowledgments vi

List of Publications vii

List of Figures xiii

Abbreviation xix

1 Introduction 1

1.1 Background 1

1.2 Research Objectives 5

1.3 Research Gap 6

1.4 Research Problem and Motivations 6

1.5 Research Significance 7

1.6 Problem Statement 7

1.7 Research Hypothesis 8

1.8 Validation 8

1.9 Research Contribution 9

1.10 Thesis Organization 9

2 Literature Survey 12

2.1 Introduction 12

2.2 The Internet of Things 12

2.3 Fog Computing* 13

2.3.1 Characterisitcs 13

2.3.2 Resource Management in Fog Model 16

2.3.3 Design and Architecture 21

2.3.4 Mobile Edge Computing 29

2.4 Data Aggregation† 31

2.4.1 Data Aggregation Reviews 32

2.4.2 Data Aggregation Models 37

2.5 group Decision Making 46

2.5.1 Group-based Recommendation Systems 46

2.5.2 GDM model 48

2.6 Consensus 49

2.7 Byzantine Fault Tolerance 54

2.8 Data Science 65

2.8.1 Data Collection 66

2.8.2 Data Analysis 67

2.8.3 Data Wrangling 68

2.8.4 Data Modelling 70

2.8.5 Train and Test 70

2.8.6 Model Evaluation 71

*parts of this section come from my paper "A review on Fog Computing technology"(2016)

†parts of this section come from my paper "A review of aggregation algorithms for the internet of things"(2017)

2.9 Machine Learning 71

2.9.1 Deep Learning 72

3 Strategy, Methodology and Computation Models 76

3.1 Introduction 76

3.2 System Topology and Operational Platform 77

3.3 Operation Phases 81

3.3.1 Pre-Exploitation Modes 81

3.3.2 Exploitation Modes 83

3.4 Consensus Management Procedure 87

3.4.1 Consensus Models 90

4 Environmental Event Detection by Consensus Bayesian Machine Learning 100

4.1 Introduction 100

4.2 Data Collection and Analysis 101

4.2.1 Pre-Expliotation Mode 102

4.3 Experimental procedure 103

4.3.1 Environment Monitoring (Normal Mode) 103

4.3.2 Event Handling (Abnormal Mode) 104

4.4 Results 110

4.5 Evaluation 114

4.6 Conclusion 118

5 Human Activity Recognition by Consensus Bayesian Deep Learning 120

5.1 Experiment description 120
5.2 Dataset 121
5.3 Assumptions 122
5.4 Data Analysis 123
5.5 Model Procedure 123
5.6 Outcome 125
5.7 Evaluation 128
5.8 Conclusion 133

6 Recognizing Smart Home Resident Activities by Consensus Bayesian Deep Learning 135

6.1 Experiment and Dataset Description 135
6.2 Assumptions 136
6.3 Data Analysis 137
6.4 Model Procedure 143
6.5 Outcomes 147
6.6 Evaluation 153
6.7 Conclusion 155

7 Conclusion 157

7.1 Introduction 157
7.2 Future Work 158

Bibliography 160

List of Figures

1.1 Mind Map 10

2.1 A typical smart home IoT scenario [1] 13

2.2 Enhanced IoT Monitoring System Scenario 14

2.3 Architecture of the Adaptive Operations Platform (AOP) 17

2.4 Comparison outcome example illustrating the improvement of the entire utility of node pairs for the proposed matching procedure [2] . 20

2.5 Interaction between clients and heterogeneous smart objects with the mediation of the IoT Hub [3] 22

2.6 Evaluating performance : a- Heap memory used (dimension in MB); b- CPU usage (dientional) [3] 24

2.7 Scenario using a normal router (a) and a router supporting DMo (b)[4] 26

2.8 Router CPU usage and the total number of packets received 27

2.9 Fog Computing Platform Yi et al. (2015)[5] 28

2.10 Case study: Co-existence of heterogeneous networks may be managed in part by clients, Chiang(2016) [6] 30

2.11 ETSIs MEC framework and reference architecture [7] 31

2.12 PBFT example 34

2.13 PBFT example results 35

2.14 Data aggregation classifications 35

2.15 Choosing responding Robot criteria. The robot at distance eleven holds an auction (red arrows used for contacting, green is for bids) in order to choose the nearest responder, which is the robot at distance five (through yellow arrows). Stojmenovic (2014)[8] 37

2.16 Flowchart of the IWD algorithm for constructing data aggregation tree in WSNs Hoang et al. (2012)[9] 40

2.17 Data aggregation process in single-hop clustered WSN Chen et al. (2008)[10] 41

2.18 Employing a tree based overlay network, like MRNet, as a scalable aggregation / analysis for real-time observation data. Bohm et al. (2010)[11] 41

2.19 Overall process of Collaborative data reduction method Park et al. (2008)[12] 42

2.20 Network model diagrams Guo et al. (2014)[13] 43

2.21 High-level system architecture and protocol.Applebaum et al. (2010)[14] 45

2.22 System model under consideration Bao and Lu (2015)[15] 46

2.23 Consensus Pattern 47

2.24 Recommendation generation process 48

2.25 Preference aggregation approach and Kemeny selection rule Muravyov and Khudonogova (2015)[16] 52

2.26 Main stages of the preference aggregation procedure Muravyov and Khudonogova (2015)[16] 53

2.27 Byzantine Generals Problem Algorithm Lamport et al. (1982) 56

2.28 Normal Case Operation of Practical Byzantine Algorithm Castro and Liskov (1999) 56

2.29 Standard WSN vs. WSN with Byzantine Algorithms Klempous (2006)[17] 59

2.30 Overview of a event stream processing system for autonomic computing Fedotova and Veltri (2006)[18] 60

2.31 BFT-MCDB Model Overview AlZain et al. (2013)[19] 62

2.32 Work Procedures of BFTCloud Zhang et al. (2011)[20] 63

2.33 Smart Metering automation scenario architecture Alaya et al. (2012)[21] 64

2.34 Data Science main Procedure 66

2.35 Training validation and test dataset and its contribution in Model prediction 71

3.1 Node Process diagram 78

3.2 IoT Fog Topology and Operational Framework 79

3.3 Process diagram 86

4.1 Environmental Temperature and Pressure sensors data distribution parameters with its corresponding resultant normalised distribution after applying the aggregation 105

4.2 Environmental Light and Humidity sensors data distribution parameters with its corresponding resultant normalised distribution after applying the aggregation 106

4.3 The difference in Packets count between the Aggregated and non-aggregated (Baseline) cases with the savings in number of packets when using different communication techniques 107

4.4 The effectiveness of Bayesian Classifier and Consensus Procedure in mode detection accuracy (3 participants nodes) 111

4.5 The effectiveness of Bayesian Classifier and Consensus Procedure in mode detection accuracy (4 participants nodes) 112

4.6 Testing the consistency of the consensus approaches on detecting the mode of operation 113

4.7 Packets savings in Aggregated versus non aggregated cases with the increase number of nodes 113

4.8 Packets savings in Aggregated versus Non Aggregated cases with the increase number of nodes for each tested transmission technique 114

4.9 consensus impact: the accuracy of the aggregated data samples increased as more nodes involved in the decision making (consensus) process 116

5.1 Assumed Topology to suit the experiment dataset into the proposed paradigm 122

5.2 Confusion Plot of the Human Activity Recognition by Consensus Bayesian Deep Learning. a-d are the four participants. 126

5.3 Confusion Plot of the Human Activity Recognition by Consensus Bayesian Deep Learning. a-c are the consensus methods, while d is the overall as if the whole data reached to the BS 127

5.4 AUC-ROC Plot of the Human Activity Recognition by Consensus Bayesian Deep Learning. a-d are the four participants 129

5.5 AUC-ROC Plot of the Human Activity Recognition by Consensus Bayesian Deep Learning. a-c are the consensus methods, while d is the overall as if the whole data reached to the BS 130

5.6 AUC-ROC Plot showing the precision of the 6 activities' result by training each of the four participants after performing the random bootstrap sampling on the data 131

5.7 AUC-ROC Plot showing the precision of the 6 activities result from the Adaboost consensus ensemble after performing the random bootstrap sampling on the data vs. the overall model 132

6.1 Assumed Topology to suit the experiment dataset into the proposed paradigm 137

6.2 Dataset Activies Frequencies showing lable inbalance 138

6.3 dataset Active vs NonActive lables 139

6.4 Training accuracy vs. validation accuracy for the dataset parts after applying the second model befor performing extensive data analysis. a-f represent the first six parts of the data (participants). 141

6.5 Accuracy for the last four parts of the dataset (participants) after applying the second model befor performing extensive data analysis (a-d). e-g are the ensembles Training vs. validation accuracies after applying the third model befor performing extensive data analysis (feature engineering) 142

6.6 comparison about prediction accuracy levels when cross-validating utilising Random forest, Logistic Regression and taking average of 10 data chuncks Bayesian Neural Network and when applying Bagging ensemble to the 10 BNN outputs for the data having one feature criterion 143

6.7 Design procedure showing the the processes done to attain Bayesian deep learning 146

6.8 prediction accuracy of the ”no activity” or ”activity” with the increas of epochs count 147

6.9 The accuracy of the first stage of distinguishing between ”Activities” classes and the ”NO Activity” class 148

6.10 uncertainity levels 149

6.11 The accuracy of participant's classifier for each of the four data subsets after applying the data analysis first and then generating the second model (a-d); e-g are the ensembles accuricies after applying the feature engineering to the dataset. 150

6.12 Confusion matrix plot showing the precision of the 12 activities' result by training each of the four participants after data analysis data analysis and performing the BNN on partitions' data 151

6.13 Confusion matrix plot showing the precision of the 12 activities' result by performing three consensus models through three different ensemble algorithms 152

6.14 Accuracies of all classes as well as the overall accuracies for the participants and ensemble algorithms after data processing and applying the BNN and consensus models 153

Abbreviation

IoT - Internet of Things

DM - Decision Making

WSN - Wireless sensor Network

DNN - Deep Neural Network

RNN - Recurrent Neural Network

CNN - Convolutional Neural Network

BNN - Bayesian Neural Network

BRNN - Bayesian Recurrent Neural Network

LSTM - Long Short-Term Memory

BS - Base Station

RPi - Raspberry Pi

AUC - Area Under the Curve

ROC - Receiver Operating Characteristics

L0 - Level zero: front end level in the proposed paradigm

L1 - Level one: Fog level in the proposed paradigm

L2 - Level two: BS or Gateway level in the proposed paradigm